

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:23:34 ; Search time 2341.82 Seconds
(without alignments)
7801.140 Million cell updates/sec

Title: US-09-788-476a-3

Perfect score: 873

Sequence: 1 tggagtgagggttaacaaga.....tgattatgtgtcttgaat 873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sy: *
12: gb_un: *
13: gb_vl: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	873	100.0	1071	6	AX048092	AX048092 Sequence
2	873	100.0	1557	9	HSAA09089	AX048092 Sequence
3	871.4	99.8	923	9	BC007099	AX048092 Sequence
4	817.8	93.7	1818	9	BC016941	AX048092 Sequence
5	647.4	74.2	84113	9	AC073063	AX048092 Sequence
6	540.4	61.9	148152	2	AC022200	AX048092 Sequence
7	540.4	61.9	149312	2	AC073328	AX048092 Sequence
8	540.4	61.9	153936	2	AP001207	AX048092 Sequence
9	540.4	61.9	166471	2	AC091052	AX048092 Sequence
10	334	38.3	408	6	AX210422	AX048092 Sequence
11	265.4	30.4	54666	6	AC073487	AX048092 Sequence
12	265.4	30.4	200379	2	AC023055	AX048092 Sequence
13	234.6	26.9	60533	2	AC105904	AX048092 Sequence
14	227.2	26.0	591	6	E25874	AX048092 Sequence
15	211.8	24.3	255	6	AX210889	AX048092 Sequence
16	90.6	10.4	156448	2	AP002493	AX048092 Sequence
17	90.6	10.4	168738	2	AP0010762	AX048092 Sequence
18	90.6	10.4	173038	2	AP001003	AX048092 Sequence
19	90.6	10.4	182429	2	AP001557	AX048092 Sequence
20	89	10.2	207874	2	AL645948	AX048092 Sequence
21	80.6	9.2	200379	2	AC023055	AX048092 Sequence
22	79.4	9.1	7218	6	166484	AX048092 Sequence
23	78.6	9.0	107794	2	AC099133	AX048092 Sequence
24	77.4	8.9	79876	2	AC097931	AX048092 Sequence
25	64.2	7.4	60533	2	AC105904	AX048092 Sequence
26	63.6	7.3	70694	3	AC006708	AX048092 Sequence
27	60	6.9	195250	2	AC091313	AX048092 Sequence
28	59.4	6.8	39816	2	AC110404	AX048092 Sequence
29	59	6.8	98894	2	AC098194	AX048092 Sequence
30	56	6.4	35875	9	AP000454	AX048092 Sequence
31	56	6.4	169328	9	AP000472	AX048092 Sequence
32	56	6.4	237569	2	AC091463	AX048092 Sequence
33	56	6.4	340000	9	AP001683	AX048092 Sequence
34	55.4	6.3	180051	23	E08995	AX048092 Sequence
35	55.4	6.3	180051	2	AC097408	AX048092 Sequence
36	55	6.3	227501	2	AC044792	AX048092 Sequence
37	54.6	6.3	205077	2	AL645807	AX048092 Sequence
38	54.4	6.2	215352	2	AC007305	AX048092 Sequence
39	54.2	6.2	152968	2	AC073669	AX048092 Sequence
40	54	6.2	5659	6	AX281376	AX048092 Sequence
41	54	6.2	5659	6	AX345277	AX048092 Sequence
42	54	6.2	5659	6	AX347367	AX048092 Sequence
43	54	6.2	5659	6	AX349088	AX048092 Sequence
44	54	6.2	85623	2	AC105726	AX048092 Sequence
45	54	6.2	169743	2	AC098448	AX048092 Sequence

ALIGNMENTS

RESULT	1	AX048092	1071 bp	DNA	Linear	PAT 15-DEC-2000
LOCUS	AX048092	Sequence	86 from Patent WO0070047.			
DEFINITION	AX048092					
ACCESSION	AX048092.1	GI:11876915				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						

Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1071)
Yang, J., Azimzal, Y., Lu, D.A., Au-Young, J. and Shih, L.L.
Full-length molecules expressed in human tissues
Patent: WO 0070047-A 86 23-NOV-2000;
Incyte Genomics, Inc. (US)
location/Qualifiers
1..1071
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 409355CB1"
BASE COUNT 338 a 190 c 297 g 246 t

ORIGIN

Query Match 100.0%: Score 873; DB 6; Length 1071;
 Best Local Similarity 100.0%: Pred. No. 6, 5e-203; Mismatches 0; Indels 0; Gaps 0;
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tggagtgagggttaacaagatgagcagcagagcgttgagctcacaataagcttgc 60
 |||||||
 DB 165 tggagtgagggttaacaagatgagcagcagcgttgagctcacaataagcttgc 224
 QY 61 cgaactaaagcagaatgtcttcctcgtggtttggaagcagaaggaataagcaagatct 120
 |||||||
 DB 225 cgaactaaagcagaatgtcttcctcgtggtttggaagcagaaggaataagcaagatct 284
 QY 121 tatccacagactccagcagcatctctgaagaacatctgaagggagcgaatgaagaaga 180
 |||||||
 DB 285 tatccacagactccagcagcatctctgaagaacatctgaagggagcgaatgaagaaga 344
 QY 181 tgaactgggagatgaacaagaggaaagaaacaaagcccatctgagctccctgtcaaga 240
 |||||||
 DB 345 tgaactgggagatgaacaagaggaaagaaacaaagcccatctgagctccctgtcaaga 404
 QY 241 ggaagaacccccctgaaaaaactgtgatgtgagcagcagagaagaagtgtgaaataatc 300
 |||||||
 DB 405 ggaagaacccccctgaaaaaactgtgatgtgagcagcagagaagaagtgtgaaataatc 464
 QY 301 atctgaatatccacagactgagagatgcagaagaagggtgaacgatccaatgtactgt 360
 |||||||
 DB 465 atctgaatatccacagactgagagatgcagaagaagggtgaacgatccaatgtactgt 524
 QY 361 gagcttgagagatgaagaagctgctcggcagcagctaggtttggagattcttcacgtccaac 420
 |||||||
 DB 525 gagcttgagagatgaagaagctgctcggcagcagcagctaggtttggagattcttcacgtccaac 584
 QY 421 aaaaagctcgtatctgataaacaacctaagttaacttgataagctggaagaaagagc 480
 |||||||
 DB 585 aaaaagctcgtatctgataaacaacctaagttaacttgataagctggaagaaagagc 644
 QY 481 tcaagaatttgcttgaatgtctctcacaatccacgaagaagctgaagatgatgaagaact 540
 |||||||
 DB 645 tcaagaatttgcttgaatgtctctcacaatccacgaagaagctgaagatgatgaagaact 704
 QY 541 gaaaaagaggaaagcagatggtggatgtgcaaaagctcaagctggaactggaaccaaga 600
 |||||||
 DB 705 gaaaaagaggaaagcagatggtggatgtgcaaaagctcaagctggaactggaaccaaga 764
 QY 601 ggaatacagaaggaaagaaagaaagaaagcagcgtttggagattgctcgtatgaagaat 660
 |||||||
 DB 765 ggaatacagaaggaaagaaagaaagaaagcagcgtttggagattgctcgtatgaagaat 824
 QY 661 tccctgatacttctgtctccagctgttccacatctctcctcctctctctgtgacacata 720
 |||||||
 DB 825 tccctgatacttctgtgttccagctgttccacatctctcctcctctctgtgacacata 884
 QY 721 tgcctcaaatgacagatcatgtgctcagctcgtcgtcgtcgaatgaggaagcatgtaccaca 780
 |||||||
 DB 885 tgcctcaaatgacagatcatgtgctcagctcgtcgtcgtcgaatgaggaagcatgtaccaca 944
 QY 781 ggtacatacgaactgcagcagcagcttgacttactgtcttgaagcttaaggttgtt 840
 |||||||
 DB 945 ggtacatacgaactgcagcagcagcttgacttactgtcttgaagcttaaggttgtt 1004
 QY 841 ggtcttctgttctgattatgttgcctgttaat 873
 |||||||
 DB 1005 ggtcttctgttctgttattgttgcctgttaat 1037

RESULT 2
 HSA409089 HSA409089 1557 bp DNA linear PRI 11-MAY-2001
 LOCUS
 DEFINITION Homo sapiens HCC-1 gene.
 ACCESSION AJ409089

VERSION AJ409089.1 GI:13940309
 KEYWORDS HCC-1 gene; hepatocellular carcinoma; nuclear protein.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1557)
 Choong,M.L., Tan,L.K., Lo,S.L., Ren,E., Ou,K., Ong,S., Liang,R.C., Seow,T.K. and Chung,M.C.
 An integrated approach in the discovery and characterization of a novel nuclear protein over-expressed in liver and pancreatic tumors
 FEBS Lett. 496 (2-3), 109-116 (2001)
 MEDLINE 21255661
 REFERENCE 2 (bases 1 to 1557)
 Choong,M.L.
 Direct Submission
 TITLE Submitted (21-FEB-2001) Choong M.L., National University of Singapore, Bioprocessing Technology Centre, MD 11 Level 5, 10 Medical Drive, 117597, SINGAPORE
 JOURNAL
 FEATURES
 source
 1..1557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7q22"
 /cell_line="HCC-M"
 /cell_type="hepatocellular carcinoma"
 /tissue_type="liver"
 1..687
 /gene="HCC-1"
 /evidence=experimental
 1..1557
 /gene="HCC-1"
 305..594
 /gene="HCC-1"
 /note="non-functional upstream open reading frame"
 /function="lightens and regulates translation of HCC-1 gene"
 446..453
 repeat_region
 /rpt_type=DIRECT
 661..1557
 /gene="HCC-1"
 688..1320
 /gene="HCC-1"
 /codon_start=1
 /product="HCC-1 protein"
 /protein_id="CAC37950.1"
 /db_xref="GI:13940310"
 /translation="MATEPVVLIHKTLAEIKOPCLARLETGKIKODLIHRLQVLE
 HAEEFANEEDVIGDETEETKPIELPKKEEPEPKTYDVAEKVKYKITSIFOTER
 MOKRAERENVPVLSKKKAAARAFGLISVPTKGLSSDNKPMVNLDKLKERQRFGLN
 VSSISRSKSEDEDEKLRKREKRGIVTSSAGTGTEDTEKKKKRAEREGIA"
 1321..1557
 /gene="HCC-1"
 /evidence=experimental
 repeat_region
 1358..1365
 /rpt_type=DIRECT
 1542
 polyA_site
 /gene="HCC-1"
 /evidence=experimental
 BASE COUNT 413 a 365 c 402 g 377 t
 ORIGIN

Query Match 100.0%: Score 873; DB 9; Length 1557;
 Best Local Similarity 100.0%: Pred. No. 6, 6e-203; Mismatches 0; Indels 0; Gaps 0;
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tggagtgagggttaacaagatgagcagcagcgttgagctcacaataagcttgc 60
 |||||||
 DB 669 tggagtgagggttaacaagatgagcagcagcgttgagctcacaataagcttgc 728
 QY 61 cgaactaaagcagaatgtcttcctcgtggtttggaagcagaaggaataagcaagatct 120

|||||
Db 729 CGAATAAGCAAGATGCTTGGCTCCGCTTTGGAGACCAAGGAATTAAGCAGATCT 788
Qy 121 tttccacagactccagagatcttttgaagaacatgcttgaagaaggaggaatgaaga 180
Db 789 TATCCACAGACTCCAGGATATCTTGAAGAATCTGAAGAAGGCAAAATGAAGA 848
Qy 181 tttactggagatgaacaagaggaagaacaacacccatctgagccctccgtcaaga 240
Db 849 TGTACTGGGATGAAACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAGA 908
Qy 241 ggaagacccccctgaaaaaactgttgatgtgagcagcagaagaagaagtgtgaataac 300
Db 909 GGAAGAACCCCTGAAAAAAGCTGTGATGTGGCAGACAGAAAGAGTGCAAAATAC 968
Qy 301 atctgaaataccacagacttgagagatgcaagaaggcctgaacgattcattccct 360
Db 969 ATCTGAATACCCACAGACTGAGAGATCCAGAAAGGGCTGACCATTCATGACCTGT 1028
Qy 361 gagcttgagagtaagaagactgtctcgagcagtaggttgagattcttcagttccaac 420
Db 1029 GAGCTTGAGAGTAAGAAAGCTGCTCGGCGACGTAAGTTTGGATTTCTTCAAGTTCCAAC 1088
Qy 421 aaaaagctgtcatctgatacaacaactatggttaacttgataagctgaagaaagac 480
Db 1089 AAAAGCTGTCTCATCTGATTAACAACCTATGTTACTTGAATAGCTGAAGAAAGAGC 1148
Qy 481 tcaaaagtgttgatgaatgtctcttcaactccagaagaagctgaagaatgataagaact 540
Db 1149 TCAAGATTTTGTGTAATGTCTCTCAATCTCAGAAAGTGTGAAGATGAGAAACT 1208
Qy 541 gaaaaagaggaagagcagatttgagattgtcacaagttcagctgaactgaacacacaga 600
Db 1209 GAAAAAGAGAGAGGAGCATTTGGGATTTGCACAAGTTCACTGGAATCGAACCACAGA 1268
Qy 601 ggatacagaggcaaaagaagaagaagcagagcgtcttgagattgcttgatgaagaagt 660
Db 1269 GGATACAGAGGCAAAAGAGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAGT 1328
Qy 661 tccgataacttctgtcttccagtggtttccatttctctctctctcttcttgtaacata 720
Db 1329 TCTGATATCTTCTGTCTCTCAAGTGTCTTCATTCCTCTCTCTTCTTGGTCAATATA 1388
Qy 721 tgcctaaatgcacagatcagtgctcagctgcctgcctgcaatgaaggagatgaaccca 780
Db 1389 TGCCTAAATGCACACTCTATGTCCTAGCTCTGCTGCATGAGAGGCAATGACCCCA 1448
Qy 781 ggtacatccatgaactgcgagcagagttgaactatgtctgttccagcttgaagttgt 840
Db 1449 GGTACATCCATGAACTGCGGCGAGCAGTTGACTTATCTGCTTTCAGCTTAAAGCTGTT 1508
Qy 841 gttgtttgttttgatgatagttgctgttaat 873
Db 1509 GTGTTTTGTTTTGATTATGTGCTGTGTTAAT 1541
RESULT 3
LOCUS BC007099 923 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, similar to RIKEN cDNA 1110005A23 gene, clone
ACCESSION BC007099
VERSION BC007099.1 GI:13937970
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 923)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer

REMARK
COMMENT
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@panix1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAL Plate: 21 Row: 9 Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
FEATURES
source
location/qualifiers
1..923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:14726 IMAGE:4273903"
/tissue_type="Prostate"
/clone_lib="NIH-MGC_83"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
22..654
/codon_start=1
/product="Similar to RIKEN cDNA 1110005A23 gene"
/protein_id="AAH07099.1"
/db_xref="GI:13937971"
/translation="MATEVELHLKILAELEKOEIARGLTEKIGIKODLIRLQAYLEE
HAEEANEEDVLGDETEEEKPIELPKVEEPEPEKTVDAAEKVKYKINSIPIOTER
MOKRAEFNVVLSKSKKAAARAFGISVPTKLSDDNKPMVLDLTKRAQRFGLN
VSSISRSSEDEKKRKRERGIYTSNAGVITTDTRAKKRRAERGIYA"
BASE COUNT 316 a 156 c 232 g 219 t
ORIGIN
Query Match 99.8%; Score 871.4; DB 9; Length 923;
Best Local Similarity 99.9%; Pred. No. 1.6e-202;
Matches 872; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 tgaagtggagggttaacaagatgagcagcagagcgttgagctccataagctaaagcttgc 60
Db 3 TGAAGTGAAGGGGTAAACAAGATGGCGACCGAGCGGTGAGCTCCATTAAGCTAAAGCTTGC 62
Qy 61 cgaactaaagcaagaatgtctgtctgtgtgttgtagagcaagaggaaataagaagatct 120
Db 63 CGAATAAGCAAGAAATCTTGTGCTCGTGTGGAGACCAAGGAATGAAGCAAGATCT 122
Qy 121 tttccacagactccagagatcttttgaagaacatgcttgaagaaggaggaatgaagaaga 180
Db 123 TATCCACAGACTCCAGGATATCTTGAAGAATCTGAAGAAGGCAAAATGAAGAAGA 182
Qy 181 tttactggagatgaacaagaggaagaagaacacccatctgagccctccgtcaaga 240
Db 183 TGTACTGGGATGAAACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAGA 242
Qy 241 ggaagacccccctgaaaaaactgttgatgtgagcagcagaagaagaagtgtgtgaataac 300
Db 243 GGAAGAACCCCTGAAAAAAGCTGTGATGTGGCAGACAGAAAGAGTGGAATAATAC 302
Qy 301 atctgaaataccacagacttgagagatgcaagaaggcctgaacgattcattccct 360
Db 303 ATCTGAATACCCACAGACTGAGAGATCCAGAAAGGGCTGACCATTCATGACCTGT 362
Qy 361 gagcttgagagtaagaagactgtctcgagcagtaggttgagattcttcagttccaac 420

Db 363 GAGCTTGAGAGTAAGAAAGCTGCTGGGACAGTAAGTTGGGATTTCTTCACCTTCAC 422

Qy 421 aaaaagctcgtcatctgatacaaacctatggttaacttgataagctgaagaaagac 480

Db 423 AAAAGTGTGTATCTGATTAACAACCTATGGTTAACTTGATTAAGCTTAAGAAAGAC 482

Qy 481 tcaaaagatttgcttgatgctcctcaatccacgaagagctgaagatgataagaaact 540

Db 483 TCAAGATTTGGTTGAATGCTCTTCATATCTCCAGAAAGTCTGAAGATGATGAGAACT 542

Qy 541 gaaaaaggaagagagagcatttgagattgtaacaaagttacagctgtaacgtgaacacaga 600

Db 543 GAAAGAGAGGAAGAGCGATTGGATTGTCAAGTTGAGCTGGAACCGAAGACACAGA 602

Qy 601 gatacagaagcaaaagaaagaaagacagacgacttgagattgctgataagaaact 660

Db 603 GGATACAGAGCGAAAGAAAGAGAAAGACAGCCCTTTGGGATTTGCTGATGAAGACT 662

Qy 661 tccgtatacttctgctcctcagtgcttccatctcctcctcctcctcctcctcctcctcct 720

Db 663 TCCGTATACTTCTGCTCTCCAGTGTTCATTTCTCTCCTCTCTCTCTCTCTCTCTCTCT 722

Qy 721 tgcctaaatgacacagctcatgctcctcagctcctcctcctcctcctcctcctcctcctcct 780

Db 723 TCCCTAAATGACAGTCAATGCTCTACGCTCCGCAATGAGGAGCATGTACCCCA 782

Qy 781 ggtacatccatgaactgcgcagcagcttgacttaattgctgcttccagcttaagtgctt 840

Db 783 GGTACATCATGATCACTGGCGACAGATTGACTTANTGCTGTTTACGCTTAAAGTTGTT 842

Qy 841 gctcttctgcttctgataatgcttctgcttctt 873

Db 843 GTGTTTGTGTTTGTATGTTGCTTGTAAAT 875

RESULT 4

LOCUS BC016941 1818 bp mRNA linear PRI 09-NOV-2001

DEFINITION Homo sapiens, clone MGC:21452 IMAGE:344846, mRNA, complete cds.

ACCESSION BC016941

VERSION BC016941.1 GI:16877383

KEYWORDS MGC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1818)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcdepaxll.stanford.edu

Dickson, M., Schmutz, J., Grilwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAC Plate: 20 Row: a Column: 13

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

Location/Qualifiers

source 1. 1818

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:21452 IMAGE:344846"

/tissue_type="Placenta, chorioarctinoma"

/clone_lib="NIH_MGC_10"

/lab_host="DH10B"

/note="Vector: pCMV-Sport6"

401. 1570

/codon_start=1

/product="unknown (protein for MGC:21452)"

/protein_id="AAH16941.1"

/db_xref="GI:16877384"

/translation="MKRAENELNSVNEFLSKLQDLLEAMNTMCSRCQKRRREFMDREPSARCAECNRLEHAEEDGFWAESMLGLITFYALMDGVYDITEMACQRYGISPTHPVPIHSPGSRIPGRGRORPADLPDLSRIEFOVPGOMPNFPAAPORPAAASRPNSTVPKRGKPKRKLALIKOSCLARETCKIKODLIHRLQAYLEHAEENEDVDLDETEBEETKRIELPVRKEEPEPTVDVAARKVKYKTSIPIOTERMQRARFNVFVLESKVARARFISVPTKGLSSDNKPMVNLKLEARQRFGLNVSISRSKSEDEKLRKRKEFIVTSAGTGTEDTEARKRKRERFGIA"

BASE COUNT 532 a 373 c 494 g 419 t

ORIGIN

Query Match 93.7%; Score 817.8; DB 9; Length 1818;

Best Local Similarity 99.8%; Pred. No. 2e-189;

Matches 819; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 53 aagcttgccgaactaaagaagaatgcttctgctgctggtttgagagcaaggaataag 112

Db 971 AAACCTGGCCGAAGTAAGCAAGAAATGCTTGTCTGCTGGTTGAGACCAAGGAATAAAG 1030

Qy 113 caagatctatccacagactccagcagcatacttgaagaacatgctgaagaggcaaat 172

Db 1031 CAAGATCTTATCCACGAGACTCCAGCATATCTTGAAGAACATGCTGAAGAGGCAAT 1090

Qy 173 gaagaagatgtaactgagagatgaacagagagaaagaaacaaagccatgaagctcct 232

Db 1091 GAAGAAGATGTACTGGAGATGAACACAGAGAAAGAAACCAAGCCCTAGCTCCCT 1150

Qy 233 gtcacaagagaagaaacccctgaaaaaactgtgtatgttgtagcagcagagaagaagtggtg 292

Db 1151 GTCAAGAGAGAAAGAACCCCTGAATAAACTGTTGATGTGCGACAGAGAAAGAGGGTG 1210

Qy 293 aaaaattacatctgaatatccacagactgagagaaatgcaagaagagtgctaaagatcaat 352

Db 1211 AAATTTACTGTGAATTAACACAGACTGAGAGATGACAAAGAGGCTTAACGATTCAAT 1270

Qy 353 gtaacctgtagcttgagagtaagaagactgctcgagcagctaggtttggaattcttca 412

Db 1271 GTAACCTGTAGCTTGAGAGTGAAGAAAGTTCGCGGACGCTTAGGTTGGATTCTTCA 1330

Qy 413 gttccaaaagagtgctgcatctgatacaaacctatggttaacttgataagctgaag 472

Db 1331 GTTCCAAACAAAGGCTGTGATCTGATTAACAAACCTATGCTTAATGGATTAAGCGAAG 1390

Qy 473 gaaagagctcaaaagatttgcttgaaatgctcctcctcaatccagaagaagctgaagatga 532

Db 1391 GAAAGAGCTCAAAAGATTGGTTGAATGCTCTTCAATCTCCAGAAAGCTGAAGATGAT 1450

Qy 533 gagaacatgaagaaagagagacgatttgagattgtaacaagctcagctgaagctga 592

Db 1451 GAGAAACTGAAGAAAGAGAGAGGACATTTGGGATTGTCAACAAGTTCACTGGAACCTGGA 1510

Qy 593 accacagagatacagaagcaagaagagaagaaagagagcagcttgtagattgctga 652

Db 1511 ACCACAGAGATACAGAGCAAGAGAAAGAGAAAGAGCAGAGGCTTGGGATTGCTTGA 1570

Qy 653 tgaagctccgatacttctgctcctcagtgcttccatctcctcctcctccttctgct 712

Db 1571 TGAAGGTTCCGATATCTTGTGTTCTCCAGTGTTTTCATTTCTCTCTCTCTCTCTGCT 1630

Qy 713 caatatatgcttaaatgacagctcatgtgctcagctcctcctcagatgagagagcat 772

```

|||||
Db 1631 CACATATATGCGCTAAATGACAGTCATGCTGCTACGTCCTCGCATGAGGAGCAT 1690
Oy 773 gtacccaggtacatccatgaactgcgcagcacgacttactatgtcgttaccgtta 832
Db 1691 GTACCCAGGTACATCCATGAGTGGGACAGAGTTTACTATTGCTGTTCACCTTTA 1750
Oy 833 aggttgctgtcttctgttcttgatctatgtctgtctta 873
Db 1751 AGGTTGTTGTCTTTTGTGTTTGTGATATGCTGCTTGAAT 1791

RESULT 5
AC073063 84113 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-136B3 from 7, complete sequence.
ACCESSION AC073063
VERSION AC073063.12 GI:15638765
KEYWORDS hrg.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 84113)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
99063792
MEDLINE 2 (bases 1 to 84113)
REFERENCE Armstrong,J., Maupin,R. and Long,N.
AUTHORS The sequence of Homo sapiens BAC clone RP11-136B3
JOURNAL Unpublished (2001)
3 (bases 1 to 84113)
REFERENCE Waterston,R.H.
AUTHORS Direct Submission
JOURNAL Submitted (08-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 84113)
REFERENCE Waterston,R.H.
AUTHORS Direct Submission
JOURNAL Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 84113)
REFERENCE Waterston,R.
AUTHORS Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 18, 2001 this sequence version replaced gi:13431102.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0136B03

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.regen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://dcpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-114P12, 2000 bp overlap; the clone sequenced to the left is GS1-259H13, 200 bp overlap. Actual start of this clone is at base position 96082 of GS1-259H13; actual end is at base position 84113 of RP11-136B3.

FEATURES
source location/Qualifiers

```

1..84113
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-136B3"
/clone_11b="RPCI-11"
22..56
/rot_family="AT-rich"
46..330
/rot_family="Alu"
335..405
/rpl_family="MERL_type"
545..613
/rpl_family="MERL_type"
614..780
/rpl_family="Alu"
781..1078
/rpl_family="Alu"
1080..1114
/rpl_family="MERL_type"
1115..1406
/rpl_family="Alu"
1407..1540
/rpl_family="MERL_type"
1833..2211
/note="match to EST BE147565 (NID:98610289)"
2322..2394
/note="match to EST BE147565 (NID:98610289)"
2337..2489
/rpl_family="Alu"
2497..2619
/rot_family="MIR"
2620..2644
/rpl_family="T)n"
2926..2991
/rpl_family="MIR"
3055..3355
/rot_family="Alu"
3374..3395
/rpl_family="T)TTTGA)n"
3552..3852
/rpl_family="Alu"
3853..3972
/rot_family="Alu"
4009..4388
/rpl_family="T2_type"
4419..4731

```

```

repeat_region /rpt_family="Alu"
4786..5085 /rpt_family="Alu"
repeat_region 5203..5238 /rpt_family="L1"
repeat_region 5239..5651 /rpt_family="L1"
repeat_region 5642..5663 /rpt_family="L1"
misc_feature 5885..6027 /note="match to EST BE795688 (NID:g10216886)"
misc_feature 5887..6025 /note="match to EST BG722929 (NID:g14002116)"
misc_feature 5887..6025 /note="match to EST BG917776 (NID:g14298252)"
misc_feature 5990..6025 /note="match to EST BG748931 (NID:g14059584)"
repeat_region 6106..6382 /rpt_family="L2"
misc_feature 6330..6334 /note="match to EST BF894562 (NID:g12286021)"
repeat_region 6383..6682 /rpt_family="Alu"
repeat_region 6683..6744 /rpt_family="L2"
repeat_region 6745..7046 /rpt_family="Alu"
repeat_region 7017..7042 /rpt_family="GAA)n"
repeat_region 7047..7139 /rpt_family="L2"
misc_feature 7251..7849 /note="match to EST AV717144 (NID:g10814296)"
repeat_region 7435..7732 /rpt_family="Alu"
repeat_region 7735..8052 /rpt_family="Alu"
misc_feature 7907..8001 /note="match to EST AV717144 (NID:g10814296)"
misc_feature 7927..8032 /note="match to EST BE147565 (NID:g8610289)"
repeat_region 8017..8086 /rpt_family="A-rich"
repeat_region 8118..8234 /rpt_family="Alu"
repeat_region 8219..8241 /rpt_family="AT-rich"
repeat_region 8245..8543 /rpt_family="Alu"
misc_feature 8300..8302 /note="match to EST AV717144 (NID:g10814296)"
repeat_region 8713..9008 /rpt_family="Alu"
repeat_region 9055..9190 /rpt_family="Alu"
repeat_region 9191..9501 /rpt_family="Alu"
repeat_region 9472..9504 /rpt_family="AT-rich"
repeat_region 9502..9674 /rpt_family="Alu"
repeat_region 9648..9686 /rpt_family="GAA)n"
repeat_region 9714..10005 /rpt_family="Alu"
repeat_region 10083..10168 /rpt_family="L1"
repeat_region 10171..10459

```

Query Match 74.2% Score 647.4; DB 9; Length 84113;
 Best Local Similarity 90.4%; Pred. No. 9.6e-148;
 Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;

```

Qy 7 gagggtaacaaatgagcagcagagtgagctccataaagctgcgcaact 66
|||||
Db 74969 GAGGGTAACAAAGATGCTGAGTGTGAGCTCCATTAAGCTGAAGCTGTAACT 75028
Qy 67 aagagaaatgctgtgctcggtgttgagacaaaggaataaagacattatcca 126
|||||
Db 75029 AAAGCAAGAAATGCTT-CTTGCTGTTTGAGACCAAGGAAATTAAGCAAGATCTTATCCA 75087
Qy 127 cagactccagcatalcttgaagaacatctgaagagagagaaatgaagaatctact 186
|||||
Db 75088 CAGACTCCAGCATATCTTGAAGAACATCTGAAGAGAGGCAAAAT--GAAGATGTACT 75144
Qy 187 gggagatgaacaagaggaagaacaagccacttgagctccctgtcaagaggaaga 246
|||||
Db 75145 AGGAGATGAACAAGAAAGAA-----AACCCCTTGCTCAATATGAGAGA 75191
Qy 247 accccctg-aaaaactgtgtgtgtgagcagagagaagaatgttgaataatcagc 305
|||||
Db 75192 ACCCCCTGAAGAAACTGTTGATGTGCGACAGAGAAAGAACTGTGAATTAATCTG 75251
Qy 306 aataccacagactgagaaatgcagaagaaggctgaacgattcaatgactgtgagct 365
|||||
Db 75252 AATACCAACAGGCTGAGAAATGCAAGAGAGGCCCAACAAATTCAGCTGTGAGCT 75311
Qy 366 tggagagtaagaagactgtctcgagcagctaggcttggagattcttcgaatccaagaag 425
|||||
Db 75312 TGGAGAGTAAAGAAAGCTGTGAGAGCTAGGTTGGAGTTCTTCAGTTCACAAAA- 75370
Qy 426 gctctcatctgataaacaactatggttaacttggataagcttgaaggaagaagctcaaa 485
|||||
Db 75371 GCTCTCATCTG--AACACACTATGGTTAACTTGATTAAGCCGAAGAAAGCTCAAA 75428
Qy 486 gatttgattgagatgtcttcattccatccagaagaatctgaagatgatgagaactgaaa 545
|||||
Db 75429 GATTGCTTGAATGCTCTTCAATTTCCAGAAAGCTCAAGATGATTAAGAAATCGAAAA 75488
Qy 546 agaagaaagagcgaattggagattgtcaaatgctcagctgaaactgaaacacagagaba 605
|||||
Db 75489 ACAGGAAGAGAGGATTTGGATTGTCCAAATTCAGCTGAACCTGAACCAACAGAGATA 75548
Qy 606 cagagagcaagaagaggaagaagcagagcgtttggagattccctgattgaagaagtcccg 665
|||||
Db 75549 CAGAGCAAAAGAGAGAAAGAGACAGAGAGCTTTGGCTTCCTGATCAAAAGTTCTGG 75608
Qy 666 atacttcgttcctccagtgatttcacatctcccttccttcctgtgcatatagctc 725
|||||
Db 75609 ATACTTCTGTTCTCCAGGTTTCCACTTCTCTCATTTCTTGGTCAATATATACGA 75668
Qy 726 aaatgcacagtcagtgtcctacgctcctgcctgcgaatgaagagacatgaccacagtac 785
|||||
Db 75669 AA-----TGCAGTCTTGGCTTGCAATGAGGAGAGGATCAACCCAGGTGA 75712
Qy 786 atccatgaactgcgacagagatttgactatgttcggttcaacttaagtggtgtgt 845
|||||
Db 75713 AACTGTGAACCTCCGCGGAGGTTGACTTATGCTTTCAGCTTTAAAGTTGTTGTT 75772
Qy 846 ttgttttgatcatgctgtcgtc 870
|||||
Db 75773 TTTGTTTTGATTATATTCCTGTT 75797

```

RESULT 6
 AC022200/c 148152 bp DNA linear HTG 09-MAY-2001
 LOCUS Homo sapiens chromosome 8 clone RP11-262017 map 8, WORKING DRAFT
 DEFINITION
 ACCESSION AC022200
 VERSION AC022200.6 GI:13518207
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

	misc_feature	/note="assembly_name:Contig32"	12957..14103	
	misc_feature	/note="assembly_name:Contig34"	14204..15942	
	misc_feature	/note="assembly_name:Contig35"	16043..17589	
	misc_feature	/note="assembly_name:Contig36"	17690..19679	
	misc_feature	/note="assembly_name:Contig37"	19780..21444	
	misc_feature	/note="assembly_name:Contig38"	21545..23225	
	misc_feature	/note="assembly_name:Contig39"	23326..25332	
	misc_feature	/note="assembly_name:Contig40"	25433..27189	
	misc_feature	/note="assembly_name:Contig41"	27290..29855	
	misc_feature	/note="assembly_name:Contig42"	29956..32915	
	misc_feature	/note="assembly_name:Contig43"	33016..35495	
	misc_feature	/note="assembly_name:Contig44"	35596..38408	
	misc_feature	/note="assembly_name:Contig45"	38509..41815	
	misc_feature	/note="assembly_name:Contig46"	41916..45472	
	misc_feature	/note="assembly_name:Contig47"	45573..49293	
	misc_feature	/note="assembly_name:Contig48"	49394..52836	
	misc_feature	/note="assembly_name:Contig49"	52997..56786	
	misc_feature	/note="assembly_name:Contig50"	clone_end:SP6 vector_side:right	
	misc_feature	/note="assembly_name:Contig51"	56887..62170	
	misc_feature	/note="assembly_name:Contig52"	62271..68265	
	misc_feature	/note="assembly_name:Contig53"	68366..72866	
	misc_feature	/note="assembly_name:Contig54"	72967..77441	
	misc_feature	/note="assembly_name:Contig55"	77542..82838	
	misc_feature	/note="assembly_name:Contig56"	82939..87223	
	misc_feature	/note="assembly_name:Contig57"	87324..93316	
	misc_feature	/note="assembly_name:Contig58"	93417..98918	
	misc_feature	/note="assembly_name:Contig59"	99019..105925	
	misc_feature	/note="assembly_name:Contig59"		
Query Match		61.9% ; Score 540.4; DB 2;	Length 149312;	
Best Local Similarity		86.9% ; Pred. No. 1.4e-121;		
Matches 64;	Conservative 0;	Mismatches 86;	Indels 11;	Gaps
OY 127	cagaatcagcgacatacttgaagaacaactgtgaaggaggaagcaaatgaagaagtact	186		
Db 134514	CTGGCTCCAGCAATATTTCCTCAGCGGCTGTGAAGAGG-----CAAATGAAGATGTACT	134567		
OY 187	gggagatgaagaacagaggaagaagaacaagc-cattgagtcacctgttaagaaggaag	245		
Db 134568	GGGAGATGAACAAGAAGAAAGAAACGCCCATTAAGAAATCCCTGTAAAGAGGAAG	134627		
OY 246	aaccccttgaaaaaaccttgatctgcagcagagaagaagtgggtgaatatcatctg	305		
Db 134628	AACCTCCGTAATAAACCTTGTTACTCTGACAGCAAGAAAGTGTGTAATAATTACATCTG	134687		
OY 306	aaataaccagagactgagagaatgcagaagaaggctgaacgatccaatcaytatcctgtgagct	365		

Db 134668	AAATACCCACAGGAGAGAGAAATGCAGAAAGGGCTGGAACAAAGTCATGTACCCGGAGCT	134747
OY 366	tggaagagtaagaagaacgtgctcggcagcctaggtttggaattcttcaagttccacaagaag	425
Db 134748	TGTAAAGTAAGAAACTGCCCGGACGCTAGGTTTGGGACCTCTTCAGTTTCAACAAAG	134807
OY 426	gtctgtcatctgtatacaaacctatggttaactctgtgataagctgaagaaagctcaaa	485
Db 134808	GTCCTGTACTCTMCAACAAACCTATCATGAACTGTGAATAGCTTAAAGGAAGAGCTCAAA	134867
OY 486	gatttggtttgaatbctctcttcacatctcagaagaatctgaagatgataagaactgaaaa	545
Db 134868	GATTTGGTTTGAATGGCTCTTCAATCTCCAGAAAGTGTGAATATGATGAGAAACTGAAAA	134927
OY 546	agaagaagaagcagatcttggtgattggttcacaaagtcagcctggaactgga---accacaag	602
Db 134928	GGAGGAAGAGACTGATTTGGGATTATCACTGTTTCAAGCTGGAAACGAAAGGAACACAGAGG	134987
OY 603	atacagaagcaagaagaagaaagacagaagcgtttggagatgctgcatgataaagttc	662
Db 134988	ATACAGAGGCAAAAGATATGAGAAAGAGCGCTCGCTATGAGGATTCGCTGTAAGATTC	135047
OY 663	ctgatacttctggtctctcgaagtttccattctctctctctctcttcttggtacacatg	722
Db 135048	ATGCGCTTTTGCTCTCCAGTGGTTTCATTTCTGTATTTCTTCGCTCACAATATA	135107
OY 723	ccataatgcacagcatgtygctctaagctcgtcctcgcgaatgagggagatgaaccag	782
Db 135108	CCTAATATGCACAGTCACTACGTACTTCTCTGCTTGCAATAGGGGACACATGTTCCCGACG	135167
OY 783	tacatccatgaacatcgcggaagcagtttgactattgctggttccagcttaagttgtgtg	842
Db 135168	TACATGTGGGAGACTCCAGACGCAATTTGACTTATGATGTTCACAC-TTAAAGCGTTGT	135226
OY 843	gttttggttttttgattatgct 862	
Db 135227	GTTTTGTTTTGGATTTT 135246	
RESULT 8		
LOCUS	AP001207 153936 bp DNA linear PRI 24-JUN-2000	
DEFINITION	Homo sapiens genomic DNA, chromosome 8q23, clone:KB1562D12.	
ACCESSION	AP001207	
VERSION	AP001207.3 GI:8698839	
KEYWORDS		
SOURCE	homo sapiens pre-pro-B cell cell_line:FLB 14 - 14 DNA, clone_lib:Keio BAC library clone:KB1562D12.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 153936) Shimizu,N. and Asakawa,S.	
JOURNAL	Homo sapiens chromosome clone BAC KB1562D12 on 8q23	
REFERENCE	Published Only in Database (2000) in press	
AUTHORS	2 (bases 1 to 153936) Shimizu,N. and Asakawa,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-FEB-2000) to the DBJ/EMBL/GenBank databases.	
COMMENT	Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)	
FEATURES	On Jun 24, 2000 this sequence version replaced gi:8096488.	
source	Location/Qualifiers	
	1..153936	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/cell_line="FLB 14 - 14"	
	/cell_type="pre-pro-B cell"	
	/chromosome="8"	
	/clone="KB1562D12"	

```
/clone_lib="keio BAC library"
/map="8q23"
repeat_region
  /rpt_family="LTR33"
  /evidence-not_experimental
    2020. .2312
  /rpt_family="AluDb"
  /evidence-not_experimental
    2403. .2462
  /rpt_family="L2"
  /evidence-not_experimental
    2795. .2827
  /rpt_family="AT-rich"
  /evidence-not_experimental
    3134. .3161
  /rpt_family="AT-rich"
  /evidence-not_experimental
    3467. .3726
  /rpt_family="MIR"
  /evidence-not_experimental
    5475. .5696
  /rpt_family="MER58A"
  /evidence-not_experimental
    6083. .6111
  /rpt_family="AT-rich"
  /evidence-not_experimental
    complement(6827. .7002)
  /rpt_family="L2"
  /evidence-not_experimental
    complement(7033. .7164)
  /rpt_family="FLAM_C"
  /evidence-not_experimental
    7203. .7308
  /rpt_family="L2"
  /evidence-not_experimental
    7433. .7546
  /rpt_family="L2"
  /evidence-not_experimental
    8592. .8715
  /rpt_family="MIR"
  /evidence-not_experimental
    8719. .8757
  /rpt_family="AT-rich"
  /evidence-not_experimental
    8770. .9035
  /rpt_family="AluJo"
  /evidence-not_experimental
    complement(10285. .10480)
  /rpt_family="MER3"
  /evidence-not_experimental
    complement(11532. .11678)
  /rpt_family="L2"
  /evidence-not_experimental
    11906. .12004
  /rpt_family="(TAA)n"
  /evidence-not_experimental
    complement(12010. .12084)
  /rpt_family="L2"
  /evidence-not_experimental
    complement(12089. .12156)
  /rpt_family="L2"
  /evidence-not_experimental
    12747. .13156
  /rpt_family="MT1J2"
  /evidence-not_experimental
    13258. .13290
  /rpt_family="AT-rich"
  /evidence-not_experimental
    13337. .13502
  /rpt_family="MIR"
  /evidence-not_experimental
    13537. .13643
  /rpt_family="L2"

repeat_region
  /evidence-not_experimental
    complement(13709. .14015)
  /rpt_family="AluJo"
  /evidence-not_experimental
    14796. .14829
  /rpt_family="(A)n"
  /evidence-not_experimental
    15290. .15526
  /rpt_family="MIR"
  /evidence-not_experimental
    15834. .15998
  /rpt_family="L1ME1"
  /evidence-not_experimental
    17379. .18211
  /rpt_family="L1ME1"
  /evidence-not_experimental
    18212. .18333
  /rpt_family="AluSx"
  /evidence-not_experimental
    18334. .18362
  /rpt_family="(CA)n"
  /evidence-not_experimental
    18363. .18517
  /rpt_family="AluSx"
  /evidence-not_experimental
    18518. .18543
  /rpt_family="(CAA)n"
  /evidence-not_experimental
    18544. .18814
  /rpt_family="L1ME1"
  /evidence-not_experimental
    complement(18991. .19309)
  /rpt_family="AluSg"
  /evidence-not_experimental
    20149. .20221
  /rpt_family="L1MC5"
  /evidence-not_experimental
    20375. .20669
  /rpt_family="AluY"
  /evidence-not_experimental
    20673. .21016
  /rpt_family="L1MC5"
  /evidence-not_experimental
    21017. .21314
  /rpt_family="AluSx"
  /evidence-not_experimental
    21315. .21613
  /rpt_family="L1MC5"
  /evidence-not_experimental
    22283. .22576
  /rpt_family="AluSp"
  /evidence-not_experimental
    24614. .24918
  /rpt_family="AluSc"
  /evidence-not_experimental
    26438. .26697
  /rpt_family="AluY"
  /evidence-not_experimental
    26893. .26919
  /rpt_family="(TTC)n"
  /evidence-not_experimental
    27106. .27295
  /rpt_family="MIR"
  /evidence-not_experimental
    27296. .27337
  /rpt_family="AT-rich"
  /evidence-not_experimental
    27464. .27948
  /rpt_family="L1MB3"
  /evidence-not_experimental
    27994. .28335
  /rpt_family="L1PA16"
  /evidence-not_experimental
```



```

Db      181 TTGGAGACTAGAAAGCTGCTCGGGACAGCTAGTTGGGATTTCTTCACTTCCACAA 240
Qy      425 ggtctcatctgtaaacacatactggttaactt-ggataagctggaagaaagct-c 482
        |||||||
Db      241 GGTCTCATCTGTAACAAACCTATGTTTACTTGGATAGCTGAAGGAAGAGCTCC 300
Qy      483 aaagatttggttaatgctctctcaatctccagaagaatc-tgaagatgataga-gaaact 540
        |||||||
Db      301 AAAGATTGGTTTGAATGTCTCTTCATCTCCAGAAAGCTCTTGAAGATGATGAGCAAACT 360
Qy      541 gaaaaaga-ggaagagagcattt-gggattgtcacaaagtcagctga 586
        |||||||
Db      361 GAAAAAGAGGAGAGAGAGCTTGGGATGTCTCAACAGTTAGCTGGA 408

RESULT 11
AC073487/c 54666 bp DNA linear PRI 03-NOV-2001
LOCUS      Homo sapiens 12 BAC RP11-76217 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION  AC073487
VERSION     AC073487.34 GI:14578058
KEYWORDS    HMG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 54666)
            Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbata,J.,
            Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Boucek,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denu,A.L., Ding,Y., Dinu,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Dubin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homosi,F.,
            Howard,S., Huber,J., Hulik,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Meshawari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,
            Morris,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nnokenkwo,S.,
            Ogunu,M., Okunou,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ken,Y., Rives,M., Rojas,A., RojudoKan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shih,C.,
            Shooshitari,N., Sisson,I., Sodergren,E., Sonaite,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,T., Telford,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wlarczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorilla,S., Zuercherlapati,R., Weinstein,G. and Gibbs,R.
            Direct Submission
TITLE      Unpublished
JOURNAL    2 (bases 1 to 54666)
AUTHORS    Worley,K.C.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (19-JUN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE  3 (bases 1 to 54666)
            Worley,K.C.
            Direct Submission
            Submitted (30-JUN-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE  4 (bases 1 to 54666)
            Worley,K.C.
            Direct Submission
            Submitted (03-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
ON Jun 30, 2001 this sequence version replaced gi:14575758.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: http://gc.bcm.tmc.edu:8088/quality_info/genbank.annotation.
html.

QUALSTAT-REPORT.
Location/Qualifiers
1. 54666
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-76217"
8. 292
/rpt_family="Alusp"
293. 320
/rpt_family="(CAAA)n"
337. 454
/rpt_family="Alub"
456. 734
/rpt_family="Alus*"
complement(1548..1602)
/rpt_family="L2"
complement(4116..4163)
/rpt_family="MIR"

```

TITLE
JOURNAL
REFERENCE
AUTHORS

repeat_region 4164..4445
/rpl_family="AluSx"
repeat_region complement(4446..4588)
/rpl_family="MIR"
repeat_region complement(4643..5004)
/rpl_family="L2"
repeat_region 5027..5086
/rpl_family="(TTTTC)n"
repeat_region complement(5096..5404)
/rpl_family="AluSx"
repeat_region complement(5738..5799)
/rpl_family="MIR"
repeat_region complement(6208..6518)
/rpl_family="AluY"
repeat_region 7220..7355
/rpl_family="C-rich"
repeat_region 7583..7801
/rpl_family="CT-rich"
repeat_region 7925..7977
/rpl_family="(CCCC)n"
repeat_region 8037..8076
/rpl_family="C-rich"
repeat_region 8212..8251
/rpl_family="(CCG)n"
repeat_region 10383..10412
/rpl_family="(GGGGA)n"
repeat_region 10522..10553
/rpl_family="(TCCC)n"
repeat_region complement(11756..11898)
/rpl_family="L2"
repeat_region 14938..15072
/rpl_family="(CAGAGA)n"
repeat_region 14968..15068
/standard_name="21494"
STS 14988..15146
/standard_name="37028"
repeat_region 15260..15281
/rpl_family="AT-rich"
repeat_region 16229..16286
/rpl_family="MIR"
STS 16447..16571
/standard_name="79124"
STS 16559..16661
/standard_name="36508"
STS 16927..17081
/standard_name="87392"
repeat_region 18299..18606
/rpl_family="AluSc"
repeat_region 18612..18788
/rpl_family="FAM"
repeat_region complement(19482..19685)
/rpl_family="MIR"
STS 21682..21805
/standard_name="73682"
STS 22161..22330
/standard_name="4986"
STS 22219..22335
/standard_name="14903"
repeat_region 23176..23469
/rpl_family="AluSg"
repeat_region 23543..23715
/rpl_family="WSTA"
repeat_region complement(23717..23966)
/rpl_family="AluSx"
repeat_region complement(24106..24219)
/rpl_family="MIR"
repeat_region complement(24220..24394)
/rpl_family="AluSx"
repeat_region 24395..24423
/rpl_family="(TAAA)n"
repeat_region complement(24424..24551)

Query Match 30.48; Score 265.4; DB 9; Length 54666;

Best Local Similarity 99.6%; Pred. No. 2,1e-54;
Matches 266; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 607 aagaggaaagagaggaagagcagagcgttggatgcctgatgaaagtcctcga 666
|
Db 22368 ACAGGCAAGAGAGGAAAAAGACAGAGCGCTTGGGATTGGCTGTAAGAGTTCTCTGA 22309
|
QY 667 taattctgtctccagtggttccattctctctctctctctctctctgtgcatatagcccta 726
|
Db 22308 TACTTTCTCTCTCCAGTCTTTTCCATTCTCTCTCTCTCTCTCTCTCTGACATATATGCCCTA 22249
|
QY 727 aatgacagtcaltgtgctacgctcctgcctgcgaatgaagagacatgtatccccaggttaca 786
|
Db 22248 AATGCACAGTCATGTGCTACGCTCCGCTCCCAATGAGGAGCATGTATCCACAGGTACA 22189
|
QY 787 tccatgaactggcggcagcaggttgacttctgtcttcttgaagtttaagtttctgtt 846
|
Db 22188 TCCATGAAGTGGCGGACGAGCTTGTGACTTATTCCTGTTTCAAGCTTAAAGGTGTGTGT 22129
|
QY 847 tctgtttgattatgttctgttctgtt 873
|
Db 22128 TTGTTTTCATATATGTGCTTGTATAT 22102
|

RESULT 12

AC023055

AC023055 200379 bp DNA linear HTG 12-JAN-2001

LOCUS

Homo sapiens chromosome 12 clone RP11-670P16, WORKING DRAFT

DEFINITION

SEQUENCE, 20 unordered pieces.

ACCESSION

AC023055

VERSION

AC023055.20 GI:12083915

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILL.

SOURCE

human.

ORANISM

Homo sapiens

REFERENCE

AUTHORS

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 200379)
Alzbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbata,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieve,M., Brown,E., Brown,M., Bryan,N.P., Bunay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carro,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,C.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,R.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Eamhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulik,S., Hume,D.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryak,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegel,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabati,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Norkenko,S.,
Ogih,M., Okunou,G., Oragunye,N., Oyedero,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitelli,N.,
Slison,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

TITLE
JOURNAL
COMMENT

Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Toppan, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (10-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: LI853
Center clone name: 320_D_22

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 726 825: gap of 100 bp
* 826 1534: contig of 709 bp in length
* 1535 1634: gap of 100 bp
* 1635 2359: contig of 725 bp in length
* 2360 2459: gap of 100 bp
* 2460 3181: contig of 722 bp in length
* 3182 3281: gap of 100 bp
* 3282 4002: contig of 721 bp in length
* 4003 4102: gap of 100 bp
* 4103 4803: contig of 701 bp in length
* 4804 4903: gap of 100 bp
* 4904 5614: contig of 711 bp in length
* 5615 5714: gap of 100 bp
* 5715 6430: contig of 716 bp in length
* 6431 6530: gap of 100 bp
* 6531 7244: contig of 714 bp in length
* 7245 7344: gap of 100 bp
* 7345 8073: contig of 729 bp in length
* 8074 8173: gap of 100 bp
* 8174 8661: contig of 688 bp in length
* 8662 8961: gap of 100 bp
* 8962 9688: contig of 727 bp in length
* 9689 9788: gap of 100 bp
* 9789 10488: contig of 700 bp in length
* 10489 10588: gap of 100 bp
* 10589 11305: contig of 717 bp in length
* 11306 11405: gap of 100 bp
* 11406 12123: contig of 718 bp in length
* 12124 12223: gap of 100 bp
* 12224 12945: contig of 722 bp in length
* 12946 13045: gap of 100 bp
* 13046 13750: contig of 705 bp in length
* 13751 13850: gap of 100 bp
* 13851 14546: contig of 696 bp in length
* 14547 14646: gap of 100 bp
* 14647 15369: contig of 723 bp in length
* 15370 15469: gap of 100 bp
* 15470 16131: contig of 662 bp in length

* 16132 16231: gap of 100 bp
* 16232 16951: contig of 720 bp in length
* 16952 17051: gap of 100 bp
* 17052 17775: contig of 724 bp in length
* 17776 17875: gap of 100 bp
* 17876 18605: contig of 730 bp in length
* 18606 18705: gap of 100 bp
* 18706 19411: contig of 706 bp in length
* 19412 19511: gap of 100 bp
* 19512 20235: contig of 724 bp in length
* 20236 20335: gap of 100 bp
* 20336 21061: contig of 726 bp in length
* 21062 21161: gap of 100 bp
* 21162 21867: contig of 706 bp in length
* 21868 21967: gap of 100 bp
* 21968 22678: contig of 711 bp in length
* 22679 22778: gap of 100 bp
* 22779 23504: contig of 726 bp in length
* 23505 23604: gap of 100 bp
* 23605 24328: contig of 724 bp in length
* 24329 24428: gap of 100 bp
* 24429 25156: contig of 728 bp in length
* 25157 25256: gap of 100 bp
* 25257 25982: contig of 726 bp in length
* 25983 26082: gap of 100 bp
* 26083 26790: contig of 708 bp in length
* 26791 26890: gap of 100 bp
* 26891 27616: contig of 726 bp in length
* 27617 27716: gap of 100 bp
* 27717 28517: contig of 801 bp in length
* 28518 28617: gap of 100 bp
* 28618 29357: contig of 740 bp in length
* 29358 29457: gap of 100 bp
* 29458 30185: contig of 728 bp in length
* 30186 30285: gap of 100 bp
* 30286 31014: contig of 729 bp in length
* 31015 31114: gap of 100 bp
* 31115 31818: contig of 704 bp in length
* 31819 31918: gap of 100 bp
* 31919 32637: contig of 719 bp in length
* 32638 32737: gap of 100 bp
* 32738 33471: contig of 734 bp in length
* 33472 33571: gap of 100 bp
* 33572 34306: contig of 735 bp in length
* 34307 34406: gap of 100 bp
* 34407 35116: contig of 710 bp in length
* 35117 35216: gap of 100 bp
* 35217 35946: contig of 730 bp in length
* 35947 36046: gap of 100 bp
* 36047 36789: contig of 743 bp in length
* 36790 36889: gap of 100 bp
* 36890 37605: contig of 716 bp in length
* 37606 37705: gap of 100 bp
* 37706 38428: contig of 723 bp in length
* 38429 38528: gap of 100 bp
* 38529 39248: contig of 720 bp in length
* 39249 39348: gap of 100 bp
* 39349 40083: contig of 735 bp in length
* 40084 40183: gap of 100 bp
* 40184 40888: contig of 705 bp in length
* 40889 40988: gap of 100 bp
* 40989 41715: contig of 727 bp in length
* 41716 41815: gap of 100 bp
* 41816 42549: contig of 734 bp in length
* 42550 42649: gap of 100 bp
* 42650 43356: contig of 707 bp in length
* 43357 43456: gap of 100 bp
* 43457 44158: contig of 703 bp in length
* 44159 44258: gap of 100 bp
* 44259 44977: contig of 719 bp in length
* 44978 45077: gap of 100 bp
* 45078 45805: contig of 728 bp in length
* 45806 45905: gap of 100 bp

*	45906	46625:	conf: of 720 bp	in length
*	46626	46725:	gap of 100 bp	in length
*	46726	47433:	conf: of 708 bp	in length
*	47434	47533:	gap of 100 bp	in length
*	47534	48269:	conf: of 736 bp	in length
*	48270	48369:	gap of 100 bp	in length
*	48370	49088:	conf: of 719 bp	in length
*	49089	49188:	gap of 100 bp	in length
*	49189	49911:	conf: of 723 bp	in length
*	49912	50011:	gap of 100 bp	in length
*	50012	50722:	conf: of 711 bp	in length
*	50723	50822:	gap of 100 bp	in length
*	50823	51533:	conf: of 711 bp	in length
*	51534	51633:	gap of 100 bp	in length
*	51634	52351:	conf: of 718 bp	in length
*	52352	52451:	gap of 100 bp	in length
*	52452	53178:	conf: of 727 bp	in length
*	53179	53278:	gap of 100 bp	in length
*	53279	54005:	conf: of 727 bp	in length
*	54006	54105:	gap of 100 bp	in length
*	54106	54822:	conf: of 717 bp	in length
*	54823	54922:	gap of 100 bp	in length
*	54923	55646:	conf: of 724 bp	in length
*	55647	55746:	gap of 100 bp	in length
*	55747	56463:	conf: of 717 bp	in length
*	56464	56563:	gap of 100 bp	in length

Matches 385; Conservative 0; Mismatches 104; Indels 38; Gaps 5;

OY	137	gcatactcttgaagaaacatgcctcgaagaaggaaggaaatgaagaagatgctactcgtggaatga	196
Db	56307	GCATATCTTTGAAGAACATGCTGCACGAAGAAGAAAAGAAATAACTATTT--GGAGATGA	56249
OY	137	acagaagagaagaacaacaagcccatctgagctccctgtcaaaagaggaagaaacccctga	256
Db	56248	ACTGAGGAAGAAGAACCAAGACGTATGAACGTGCTGTGAAGAAGAGAA--CACTGA	56191
OY	257	aaaactgtctgattgca-----gcagagaagaa	286
Db	56190	AAGCTGGTGTATGCAATGCATCACACACACACACACACACACACACACACACACACAAA	56131
OY	287	gttgtgaaatcacatctgaaataccacagactgagaatgcaagaagagctgaacga	346
Db	56130	GTGTAATAAATTTACATCTGGAATATCTCAGACTGAGAAATACAGAA--ATTGAATGA	56074
OY	347	ttcaatgtaacttbgagcttggaggtgaagaagctgtcggcagctagcttggatt	406
Db	56073	GTCATATGTAACCTGAAAGTTTGAGAGACAAGAGGCTGCTCAAGTGATGAGTTTGAAAT	56014
OY	407	tcttcagttccaacaagaagctcgtgcacatctgaataacaacatatgttaacttggataag	466
Db	56013	TCTTCAGTTCCCAATFAGAGGTTTCATCATCTGTGCACCAAGCTTATAGTTTAAGCTGGAAAA	55954
OY	467	ctgaagagaagaagctccaagaatttggttgaatgctcttcaatctccaagaagctctga	526
Db	55953	TTAAAGGAAAGACACATAGATTTGGTTTGAAATGCTCTTCCAAAACCTAGAAATTTCTGAG	55894
OY	527	gactatggaacactbaaaaaagagaagagcgacttggattgtgtcaacagtcaactgca	586
Db	55893	GATCATGAGAGGCTCAAAAATCAGAA--AAGATTTGGGATTTTGACCAAGTTTAGGTGA	55836
OY	587	actggaaccacagaggaatacagagggcacaagaagagaaagaagacga 633	
Db	55835	AATGAAAAACCCAGAGGACAGAGAGCAATATAAATAAAAAACAAAACA 55789	

RESULT	14				
LOCUS	E25874				
DEFINITION	E25874				
ACCESSION	Blastocyst	CDNA.			
	E25874				
		591 bp	DNA	1 linear	PAT 07-FEB-2001

VERSION	E25874.1	GI:13026054
KEYWORDS	JP 1999164691-A/3.	
SOURCE	unidentified.	
ORGANISM	unidentified.	
REFERENCE	1 (bases 1 to 591)	
AUTHORS	Yoshinide, H.M.K.K.	
TITLE	Blastocyst cDNA	
JOURNAL	Patent: JP 1999164691-A 3 22-JUN-1999;	
COMMENT	RIKAGAKU KENKYUSHO	
OS	Unidentified	
PN	JP 1999164691-A/3	
PD	22-JUN-1999	
PF	14-APR-1998	JP 1998103115
PR		
PI	YOSHITIDE HAYASHIZAKI, MORIYAKI KUSAKABE	
PC	Cl2N15/09, Cl2O1/68, Cl2N15/00	
CC	Strandedness: Single;	
CC	Topology: Linear;	
CC	Key	
FT	Location/Qualifiers	
FT	source	1..591
FEATURES	Location/Qualifiers	
source	1..591	
	/organism="unidentified"	
	/db_xref="taxon:32644"	
BASE COUNT	177 a	82 c 139 g 172 t 21 others
ORIGIN		

Query Match	26.0%;	Score 227.2;	DB 6;	Length 591;
Best Local Similarity	69.3%;	Pred. No. 4.1e-45;		
Matches 389;	Conservative 0;	Mismatches 147;	Indels 25;	Gaps 6;

Oy	318	ctagaagatgcgaagaaggggcctgcagatcaatgtaacgtctcgttgcgcttgaggaataaga	37
Db	18	CAGAAATGANAAGACGAGAGGGGCTGAAGTTTAATAAATCTTAAGC--TGAGAGATAGA	76
Oy	378	aagctgcctgcgcagctaggttggttgattctctcagttccacaacaaagctcgtcatcty	43
Db	77	AGGCTGCCCG--NCGTTAGNTTGGAAATTTTTCAGTTTCCAGTTCCAACAANAGTTTTTNN---TG	13
Oy	438	atacaaacactatggtttaacttgcgataagctggaagaaagagctcaagatttgattga	49
Db	132	GACACNACCCNNTGGTTTAACNTGNTTAANTAGGGGAAAGGCGACACAGAGA--TGGGGNNMG	19
Oy	498	atgctcttcaatctccagaagaatcttgaagatgatgagaacatggaacaaagagaagcgc	55
Db	191	ATNTTCTTCCATNTCCNAGAAAGCTGTGGGATGATGAGAGAGCTGGAAGAAACCGAAGGAGA	25
Oy	558	gattcggagattgcacaagtgcagctgcgaaactgcggaaccaagaagatacagaaggcaaga	61
Db	251	GATTNGGATTTGTGCACAACTTCAGNTTGAATGCAACCGACAGACGATACAGAGCCAAAGA	31
Oy	618	agaagaaaagcagagcgccttgcgattgcctcgtatgaaagtctctgatacttcgtt	67
Db	311	AAAGAAAAGACGACAGAGCGTTTGGAAATGCATAATGAAAGTGTGTG---CTTCTCGA	36
Oy	678	ctcagtgattccattctctctctctctctcttggtcacaatatatgcttaaatgcagtc	73
Db	368	CCATGATGTTTCCATCTCTCAAAATTTTGTTTATATATCATACACACCCCTAC	42
Oy	738	at-----gtgcctacgctcgtgcctgcgacatagaggagatgtaaccacag	78
Db	428	CTACCTACCTATTTCATACACACAGTCATGTGCTTACATGTGGAAAGCTGTACCTTTGG	48
Oy	783	tacatccatgaactgcgcgcagcagcttgcatactatcgcgtttcagaacttgaagttcgt	84
Db	488	TACATCCTGTAATTCAGTAGCAGTTTATTTAAATGCGTTTNTGGGTTTAACATTTGTTGT	54
Oy	843	gttttctgttttgattgattgct 863	
Db	548	GGTTTTGTGTTTAATTAATT 568	

```
RESULT 15
AX210889 255 bp DNA linear PAT 31-AUG-2001
LOCUS DEFINITION Sequence 531 from Patent WO0157058.
ACCESSION AX210889
VERSION AX210889.1 GI:15425149
KEYWORDS
SOURCE
ORGANISM
Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 255)
AUTHORS Rosenthal,A., Hinzmann,B., Schaefer,R., Zuber,J., Tchernitsa,O.,
Grips,M., Hellriegel,M., Schmitz,A.C. and Sers,C.
TITLE Detection of differential gene expression
JOURNAL Patent: WO 0157058-A 531 09-AUG-2001;
Metagen Gesellschaft fuer Genomforschung mbH (DE)
FEATURES
source 1..255
/organism="Rattus sp."
/db_xref="taxon:10118"
BASE COUNT 89 a 40 c 69 g 57 t
ORIGIN

Query Match 24.3%; Score 211.8; DB 6; Length 255;
Best Local Similarity 89.4%; Pred. No. 2.3e-41;
Matches 228; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 184 actggagatgaaacagagaagaacaaagccattgagctccctgtcacaagaga 243
Db 1 ACTGGAGATGAAGCTGAGAGAGAAACCAAGCCTATAGAACTGCCGTTAAAGAGGA 60
QY 244 agaacccctgaaaaaactgtgtgattggtcagcagagaagaagtgtgaaaaattacatc 303
Db 61 AGAACCTCTCTGAAAAAGTTGTGATATGGCATCAGAAAAAGGTGTAAATAATTACATC 120
QY 304 tgaattaccacagacttgagagaatgcagaagaaggctgaacgatccaatgtacactgtgag 363
Db 121 TCGAATACCTCAACTGAGAGAAATCCAGAAAGGCGCTGAACTTCAATGTGCTGTAAAG 180
QY 364 ctggagaagtaagaagctgtcggcagctaggtttggatttcctcagttccacaaca 423
Db 181 CTTGAGAGAGTAAGAAGCTGCTCGGCGAGCGAGTTTGAATTTCTTCAGTTCCAAACAA 240
QY 424 aggtctgtcatctga 438
Db 241 AGTTTATCATCTGA 255
```

Search completed: June 19, 2002, 14:32:08
Job time: 9541 sec